

How glassy are neural networks?

Adriano Barra,^{*} Giuseppe Genovese,[†] Francesco Guerra,[‡] Daniele Tantari,[§]

March 4, 2013

Abstract

In this paper we continue our investigation on the high storage regime of a neural network with Gaussian patterns. Through an exact mapping between its partition function and one of a bipartite spin glass (whose parties consist of Ising and Gaussian spins respectively), we give a complete control of the whole annealed region. The strategy explored is based on an interpolation between the bipartite system and two independent spin glasses built respectively by dichotomic and Gaussian spins: Critical line, behavior of the principal thermodynamic observables and their fluctuations as well as overlap fluctuations are obtained and discussed. Then, we move further, extending such an equivalence beyond the critical line, to explore the broken ergodicity phase under the assumption of replica symmetry and we show that the quenched free energy of this (analogical) Hopfield model can be described as a linear combination of the two quenched spin-glass free energies even in the replica symmetric framework.

Introduction

Neural networks, thought of as the *harmonic oscillators* of artificial intelligence, are nowadays being used in a huge number of different fields of science, ranging from practical application in data mining [13, 39] to theoretical speculation in systems biology [2, 22], crossing fields as disparate as computer science [34], quantitative sociology [8] or economics [19].

As a consequence, as applications develop, the need for mathematical methods (bringing them under rigorous control) and a simple mathematical framework (acting as a benchmark for future speculation) increases and motivates

^{*}Dipartimento di Fisica, Sapienza Università di Roma and GNFM, Sezione di Roma.

[†]Dipartimento di Matematica, Sapienza Università di Roma.

[‡]Dipartimento di Fisica, Sapienza Università di Roma and INFN, Sezione di Roma.

[§]Dipartimento di Matematica, Sapienza Università di Roma.

the present paper.

Moreover, although the Hopfield model has been extensively studied since it was introduced in [33], both from a physical [5, 6, 14, 21, 24] and a more mathematical [4, 15, 16, 17, 36, 37, 42, 43] point of view, from the rigorous perspective many points about its properties remain unsolved, which also prompts further efforts in developing new mathematical techniques and different physical perspectives.

In the past, we gave an extensive treatment of an analogical neural network [9][10], namely a mean-field structure with N dichotomic neurons (spins) interconnected through Hebbian couplings [5, 21] whose p patterns are stored according to a standard Gaussian $\mathcal{N}[0, 1]$: In [9] we studied its thermodynamical properties paying attention to the annealed approximation (but we were unable, at that time, to gain a complete control of the whole annealed region), while in [10] we investigated the properties of the replica symmetric approximation.

Within our approach, the equilibrium statistical mechanics of the neural network is shown to be equivalent to the one of a bipartite spin glass whose parts consist of the original N neurons (belonging to the first party, hence made of by dichotomic variables) and the other hand p Gaussians that give rise to the second part (hence consisting of continuous variables): As the theory of the mean-field Ising spin glass (namely the Sherrington Kirkpatrick model [35]) has been intensively developed in the past decades (see for instance [3][7][28][31]), while the same did not happen for the Gaussian counterpart, we investigated in detail the structure of the latter too, deepening the understanding of its properties in [11].

Furthermore, to complete a streamlined description of the state of the art on this theme, we stress that results on the analogical Hopfield model, stemming from a mathematical perspective far from our connection with bipartite spin glasses, have also been obtained in [15, 16, 17].

Turning to the applied side, despite the fact that in neural networks (in their original artificial intelligence framework) the interest in continuous patterns is reduced or moved to rotators (e.g. Kuramoto oscillators [1]), as digital processing by Ising spins works as a better approximation for the standard *integrate and fire* models of neurons [18], in several other fields of science (as, for instance, in chemical kinetics [22, 23] or theoretical immunology [2]) continuous values of patterns can instead be preferred ([14][20]) and a rigorous mathematical control of completely continuous models (namely with both continuous patterns and neurons) belongs to our strategy of research. For the moment, we limit ourselves in presenting a clear scenario for the hybrid model made of by continuous patterns and dichotomic variables, namely

the analogical neural network: In Section One we introduce the model and all the statistical-mechanics-related concepts. Then in Section Two we expose our new strategy of interpolation which allows a complete control both of the ergodic region (confirming the annealed approximation, which is investigated in great detail), and of the replica symmetric scenario, which is then deepened in Section Three.

The last section contains our conclusions.

Furthermore, an appendix is added: there the fluctuation theory of the order parameters of the model is discussed, and it is shown that the critical line found in this work characterizes a second order phase transition.

1 The model, basic definitions and properties

1.1 The analogical Hopfield model

We introduce a large network of N two-state neurons $(1, \dots, N) \ni i \rightarrow \sigma_i = \pm 1$, which are thought of as quiescent when their value is -1 or spiking when their value is $+1$. They interact throughout a synaptic matrix J_{ij} defined according to the Hebb rule for learning [32, 33]

$$J_{ij} = \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu. \quad (1)$$

Each random variable $\xi^\mu = \{\xi_1^\mu, \dots, \xi_N^\mu\}$ represents a learned pattern: While in the standard literature these patterns are usually chosen at random independently with values ± 1 taken with equal probability $1/2$, we chose them as taking real values with a unit Gaussian probability distribution, *i.e.*

$$d\mu(\xi_i^\mu) = \frac{1}{\sqrt{2\pi}} e^{-(\xi_i^\mu)^2/2}. \quad (2)$$

The analysis of the network assumes that the system has already stored p patterns (no learning is investigated here), and we will be interested in the case in which this number asymptotically increases linearly with respect to the system size (high storage level), so that $p/N \rightarrow \alpha$ as $N \rightarrow \infty$, where $\alpha \geq 0$ is a parameter of the theory denoting the storage level.

The Hamiltonian of the model has a mean-field structure and involves interactions between any pair of sites according to the definition

$$H_N(\sigma; \xi) = -\frac{1}{N} \sum_{\mu=1}^p \sum_{i < j}^N \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j. \quad (3)$$

1.2 Morphism in the bipartite model

By splitting the summations $\sum_{i<j}^N = \frac{1}{2} \sum_{ij}^N - \frac{1}{2} \sum_i^N \delta_{ij}$ in the Hamiltonian (3), we can introduce and write the partition function $Z_{N,p}(\beta; \xi)$ in the following form

$$\begin{aligned} Z_{N,p}(\beta; \xi) &= \sum_{\sigma} \exp \left(\frac{\beta}{2N} \sum_{\mu=1}^p \sum_{ij}^N \xi_i^{\mu} \xi_j^{\mu} \sigma_i \sigma_j - \frac{\beta}{2N} \sum_{\mu=1}^p \sum_i^N (\xi_i^{\mu})^2 \right) \quad (4) \\ &= \tilde{Z}_{N,p}(\beta; \xi) e^{\frac{-\beta}{2N} \sum_{\mu=1}^p \sum_{i=1}^N (\xi_i^{\mu})^2} \end{aligned}$$

where $\beta \geq 0$ is the inverse temperature, and denotes here the level of noise in the network. We have defined

$$\tilde{Z}_{N,p}(\beta; \xi) = \sum_{\sigma} \exp \left(\frac{\beta}{2N} \sum_{\mu=1}^p \sum_{ij}^N \xi_i^{\mu} \xi_j^{\mu} \sigma_i \sigma_j \right). \quad (5)$$

Notice that the last term at the r.h.s. of eq. (4) does not depend on the particular state of the network, hence the control of the last term can be easily obtained [9] and simply adds a factor $\alpha\beta/2$ to the free energy.

Consequently we focus just on $\tilde{Z}(\beta; \xi)$. Let us apply the Hubbard-Stratonovich lemma [25] to linearize with respect to the bilinear quenched memories carried by the $\xi_i^{\mu} \xi_j^{\mu}$.

We can write

$$\tilde{Z}_{N,p}(\beta; \xi) = \sum_{\sigma} \int \left(\prod_{\mu=1}^p \frac{dz_{\mu} \exp(-z_{\mu}^2/2)}{\sqrt{2\pi}} \right) \exp \left(\sqrt{\beta/N} \sum_{i,\mu} \xi_i^{\mu} \sigma_i z_{\mu} \right). \quad (6)$$

For a generic function F of the neurons, we define the Boltzmann state $\omega_{\beta}(F)$ at a given level of noise β as the average

$$\omega_{\beta}(F) = \omega(F) = (Z_{N,p}(\beta; \xi))^{-1} \sum_{\sigma} F(\sigma) e^{-\beta H_N(\sigma; \xi)} \quad (7)$$

and often we will drop the subscript β for the sake of simplicity. The s -replicated Boltzmann state is defined as the product state $\Omega = \omega^1 \times \omega^2 \times \dots \times \omega^s$, in which all the single Boltzmann states are at the same noise level β^{-1} and share an identical sample of quenched memories ξ . For the sake of clearness, given a function F of the neurons of the s replicas and using the symbol $a \in [1, \dots, s]$ to label replicas, such an average can be written as

$$\Omega(F(\sigma^1, \dots, \sigma^s)) = \frac{1}{Z_{N,p}^s} \sum_{\sigma^1} \sum_{\sigma^2} \dots \sum_{\sigma^s} F(\sigma^1, \dots, \sigma^s) \exp \left(-\beta \sum_{a=1}^s H_N(\sigma^a, \xi) \right). \quad (8)$$

The average over the quenched memories will be denoted by \mathbb{E} and for a generic function of these memories $F(\xi)$ can be written as

$$\mathbb{E}[F(\xi)] = \int \left(\prod_{\mu=1}^p \prod_{i=1}^N \frac{d\xi_i^\mu}{\sqrt{2\pi}} e^{-\frac{(\xi_i^\mu)^2}{2}} \right) F(\xi) = \int F(\xi) d\mu(\xi), \quad (9)$$

with $\mathbb{E}[\xi_i^\mu] = 0$ and $\mathbb{E}[(\xi_i^\mu)^2] = 1$.

Hereafter we will often denote the average over the gaussian spins as $d\mu(z)$.

We use the symbol $\langle . \rangle$ to mean $\langle . \rangle = \mathbb{E}\Omega(.)$.

We recall that in the thermodynamic limit it is assumed

$$\lim_{N \rightarrow \infty} \frac{p}{N} = \alpha,$$

α being a given real number, which acts as free parameter of the theory.

1.3 The thermodynamical observables

The main quantities of interest are the intensive pressure, defined as

$$\lim_{N \rightarrow \infty} A_{N,p}(\beta, \xi) = -\beta \lim_{N \rightarrow \infty} f_{N,p}(\beta, \xi) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln Z_{N,p}(\beta; \xi), \quad (10)$$

the quenched intensive pressure, defined as

$$\lim_{N \rightarrow \infty} A_{N,p}^*(\beta) = -\beta \lim_{N \rightarrow \infty} f_{N,p}^*(\beta) = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \ln Z_{N,p}(\beta; \xi), \quad (11)$$

and the annealed intensive pressure, defined as

$$\lim_{N \rightarrow \infty} \bar{A}_{N,p}(\beta) = -\beta \lim_{N \rightarrow \infty} \bar{f}_{N,p}(\beta) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \mathbb{E} Z_{N,p}(\beta; \xi). \quad (12)$$

According to thermodynamics, here $f_{N,p}(\beta, \xi) = u_{N,p}(\beta, \xi) - \beta^{-1} s_{N,p}(\beta, \xi)$ is the free energy density, $u_{N,p}(\beta, \xi)$ is the internal energy density and $s_{N,p}(\beta, \xi)$ is the intensive entropy (the star and the bar denote the quenched and the annealed evaluations as well).

According to the exploited bipartite nature of the Hopfield model, we introduce two other order parameters: the first is the overlap between the replicated neurons, defined as

$$q_{ab} = \frac{1}{N} \sum_{i=1}^N \sigma_i^a \sigma_i^b \in [-1, +1], \quad (13)$$

and the second the overlap between the replicated Gaussian variables z , defined as

$$p_{ab} = \frac{1}{p} \sum_{\mu=1}^p z_{\mu}^a z_{\mu}^b \in (-\infty, +\infty). \quad (14)$$

These overlaps play a considerable role in the theory as they can express thermodynamical quantities.

2 A detailed description of the annealed region

2.1 The interpolation scheme for the annealing

In this section we present the main idea of the work, used here to get a complete control of the high-temperature region: We interpolate between the neural network (described in terms of a bipartite spin glass) and a system consisting of two separate spin glasses, one dichotomic and one Gaussian. Note that, by the Jensen inequality, namely

$$\mathbb{E} \ln Z_{N,p}(\beta) \leq \ln \mathbb{E} Z_{N,p}(\beta),$$

we can write

$$A_{N,p}^* \leq \frac{1}{N} \ln \mathbb{E} \sum_{\sigma} \int \prod_{\mu=1}^p d\mu(z_{\mu}) e^{\sqrt{\frac{\beta}{N}} \sum_{i\mu} \xi_i^{\mu} \sigma_i z_{\mu}} = \ln 2 - \frac{p}{2N} \log(1 - \beta), \quad (15)$$

where we emphasize that the integral inside eq. (15) exists only for $\beta < 1$. The $N \rightarrow \infty$ limit then offers immediately $\lim_{N \rightarrow \infty} A_{N,p}^*(\beta) \leq \ln 2 - \alpha \ln(1 - \beta)/2$. The next step is to use interpolation to prove the validity of the Jensen bound in the whole region defined by the line $\beta_c = 1/(1 + \sqrt{\alpha})$, which defines the boundary of the validity of the annealed approximation, in complete agreement with the well known picture of Amit, Gutfreund and Sompolinsky [5][6].

To understand which is the proper interpolating structure, let us note that the exponent of the Boltzmann factor yields a family of random variables indexed by the configurations (σ, z) . For a given realization of the noise, $H(\sigma, z|\xi) = \sqrt{\frac{\beta}{N}} \sum_{i\mu} \xi_{i,\mu} \sigma_i z_{\mu}$ is a randomly centered variable with variance

$$\mathbb{E}(H(\sigma, z|\xi) H(\sigma', z'|\xi)) = \frac{\beta}{N} \sum_{i\mu} \sigma_i \sigma'_i z_{\mu} z'_{\mu} = \beta p q_{\sigma\sigma'} p_{zz'}.$$

The presence of the product $q_{\sigma\sigma'}p_{zz'}$ in the variance suggests the correct interpolating structure among this bipartite network and two other independent spin glasses, namely a Sherrington-Kirkpatrick model with variance $q_{\sigma\sigma'}^2$ and another spin glass model with Gaussian spin and variance $p_{zz'}^2$. It is in fact clear that a proper interpolating structure can be held by

$$\begin{aligned}\varphi_N(t) &= \frac{1}{N} \mathbb{E} \ln \sum_{\sigma} \int \prod_{\mu=1}^p d\mu(z_{\mu}) \exp \left(\sqrt{t} \sqrt{\frac{\beta}{N}} \sum_{i\mu} \xi_i^{\mu} \sigma_i z_{\mu} \right) \quad (16) \\ &\cdot \exp \left(\sqrt{1-t} \left(\beta_1 \sqrt{\frac{N}{2}} K(\sigma) + \beta_2 \sqrt{\frac{p}{2}} \bar{K}(z) \right) \right) \\ &\cdot \exp \left((1-t) \left(\frac{p\beta}{2} p_{zz} - \frac{p\beta_2^2}{4} p_{zz}^2 \right) \right),\end{aligned}$$

where we have set

$$K(\sigma) = \frac{1}{N} \sum_{ij} J_{ij} \sigma_i \sigma_j$$

and

$$\bar{K}(z) = \frac{1}{p} \sum_{ij} \bar{J}_{ij} z_i z_j$$

and the average \mathbb{E} is taken with respect to all the i.i.d. normal random variables $\xi_{ij}, J_{ij}, \bar{J}_{ij}$. The interpolation is performed such that for $t = 1$ the interpolating structure $\varphi(t = 1)$ returns the free energy of the bipartite model, namely of the neural network, while for $t = 0$ it coincides with a factorization in an SK spin glass and a (suitably regularized) Gaussian one [11]; β_1, β_2 , which will be then fixed as opportune noise levels, for the moment are simply free parameters.

As in [10][30], the plan is now to evaluate the flow under a changing t of the interpolating structure in order to get a positive defined sum rule by tuning opportunely β_1, β_2 ; hence, if we generalize the states as $\langle \cdot \rangle_t = \mathbb{E} \Omega_t$, where the subscript t accounts for the extended interpolating structure defined in (16) we can write

$$\frac{d\varphi_N(t)}{dt} = \frac{1}{N} \frac{1}{2} \beta p \left(\langle p_{zz} \rangle_t - \langle q_{\sigma\sigma'} p_{zz'} \rangle_t \right) - \frac{1}{4} \beta_1^2 \left(1 - \langle q_{\sigma\sigma'}^2 \rangle_t \right) + \quad (17)$$

$$- \frac{p}{N} \frac{1}{4} \beta_2^2 \left(\langle p_{zz}^2 \rangle_t - \langle p_{zz'}^2 \rangle_t \right) + \frac{p}{N} \frac{1}{4} \beta_2^2 \langle p_{zz}^2 \rangle_t - \frac{\beta}{2} \frac{p}{N} \langle p_{zz} \rangle_t, \quad (18)$$

then, calling $\alpha = p/N$ even at finite size N (with a little language abuse), we can write

$$\frac{d\varphi_N(t)}{dt} = -\frac{\beta_1^2}{4} + \frac{1}{4} \langle \beta_1^2 q_{\sigma\sigma'}^2 + \alpha \beta_2^2 p_{zz'}^2 - 2\alpha \beta q_{\sigma\sigma'} p_{zz'} \rangle_t. \quad (19)$$

If we now impose on β_1, β_2 the constraint $\beta_1\beta_2 = \sqrt{\alpha}\beta$ we get a perfect square in the brackets of the flow under a changing t , and calling $S_t(\alpha, \beta) = \langle (\beta_1 q_{\sigma\sigma'} - \sqrt{\alpha}\beta_2 p_{zz'})^2 \rangle_t$ the source term, we can write

$$\frac{d\varphi_N}{dt} \geq -\frac{1}{4}\beta_1^2 + S_t(\alpha, \beta). \quad (20)$$

We can then integrate back between $[0, 1]$ to get the following inequality

$$\begin{aligned} \varphi_N(1) &= \frac{1}{N} \mathbb{E} \ln \sum_{\sigma} \int \prod_{\mu}^p d\mu(z_{\mu}) e^{\sqrt{\frac{\beta}{N}} \sum_{i\mu} \xi_i^{\mu} \sigma_i z_{\mu}} \\ &\geq \frac{1}{N} \mathbb{E} \ln \sum_{\sigma} e^{\beta_1 \sqrt{\frac{N}{2}} K(\sigma)} \\ &\quad - \frac{\beta_1^2}{4} + \frac{p}{N} \frac{1}{p} \mathbb{E} \ln \int \prod_{\mu} d\mu(z_{\mu}) e^{\beta_2 \sqrt{\frac{p}{2}} \bar{K}(z)} e^{-\frac{\beta_2^2 p}{4} p_{zz'}} e^{\frac{p}{2} \beta p_{zz}}, \end{aligned}$$

under the constraint $\beta_1\beta_2 = \sqrt{\alpha}\beta$.

Note that $K(\sigma)$ in the above expression defines the SK-model, while the last term defines the regularized Gaussian spin glass deeply investigated in [11]. Now the advantages of this interpolation scheme become evident: As we have extremely satisfactory descriptions of the two independent models, namely the SK and the Gaussian spin glass, by these properties we can infer the behavior of the neural network (again thought of as the bipartite spin glass). In particular, we know that the free energies of each single part spin glass approach their annealed expression in the region where $\beta_1 \leq 1$ [41] and $\beta + \beta_2 \leq 1$ [11]. Within this region, at the r.h.s. of eq. (21) we get, in the thermodynamic limit, exactly $\ln 2 - (\alpha/2) \ln(1 - \beta)$.

Furthermore, if α and β respect the constraint $\beta(1 + \sqrt{\alpha}) \leq 1$, then finding β_1, β_2 such that the conditions (A), (B), (C) hold, being

$$\beta_1\beta_2 = \sqrt{\alpha}\beta \quad (A), \quad \beta_1 \leq 1 \quad (B), \quad \beta + \beta_1 \leq 1 \quad (C),$$

is certainly possible. In particular, using the SK critical behavior for the sake of simplicity, hence posing $\beta_1 = 1$, and setting $\beta_2 = \sqrt{\alpha}\beta$, conditions (A) and (B) are automatically satisfied and, for the latter, being $\beta_2 = \sqrt{\alpha}\beta$, we get

$$\beta + \beta_2 \equiv \beta + \sqrt{\alpha}\beta = \beta(1 + \sqrt{\alpha}) \leq 1,$$

such that also condition (C) is verified. We can then state the following

Theorem 1. *In the α, β plane there exist a critical line, defined by*

$$\beta_c(\alpha) = \frac{1}{1 + \sqrt{\alpha}}, \quad (21)$$

such that for $\beta \leq \beta_c(\alpha)$ the annealed approximation of the free energy holds

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \ln \sum_{\sigma} \int \prod_{\mu} d\mu(z_{\mu}) e^{\left(\sqrt{\frac{\beta}{N}} \sum_{i\mu} \xi_i^{\mu} \sigma_i z_{\mu} \right)} = \ln 2 - \frac{\alpha}{2} \ln(1 - \beta). \quad (22)$$

Remark 1. *We stress that the Borel-Cantelli lemma allows straightforwardly to determine the correct annealed regions for the SK model [41] and, through a careful check of convergence of the integral defining the partition function, the same holds for the Gaussian case too [11]; however, the direct application of the Borel-Cantelli argument on the neural network gives a weaker result as shown for instance in [9]. The interpolation scheme allows to exploit and transfer the results for the SK and Gaussian models to the neural network, and enlarges the area of validity of the annealed expression for the free energy to the whole expected region, obtained e.g. via the replica method [5].*

2.2 The control of the annealed region

As a consequence, we can now extend the previous results exposed in [9] to the whole annealed region: Summarizing, we get the following

Theorem 2. *There exists $\beta_c(\alpha)$, defined by eq. (21), such that for $\beta < \beta_c(\alpha)$ we have the following limits for the intensive free energy, internal energy and entropy, as $N \rightarrow \infty$ and $p/N \rightarrow \alpha > 0$:*

$$\begin{aligned} -\beta \lim_{N \rightarrow \infty} f_{N,p}(\beta; \xi) &= \lim_{N \rightarrow \infty} N^{-1} \ln Z_{N,p}(\beta; \xi) \\ &= \ln 2 - (\alpha/2) \ln(1 - \beta) - (\alpha\beta/2), \end{aligned} \quad (23)$$

$$\begin{aligned} \lim_{N \rightarrow \infty} u_{N,p}(\beta; \xi) &= - \lim_{N \rightarrow \infty} N^{-1} \partial_{\beta} \ln Z_{N,p}(\beta; \xi) \\ &= -\alpha\beta/(2(1 - \beta)), \end{aligned} \quad (24)$$

$$\begin{aligned} \lim_{N \rightarrow \infty} s_{N,p}(\beta; \xi) &= \lim_{N \rightarrow \infty} N^{-1} (\ln Z_{N,p}(\beta; \xi) - \beta \partial_{\beta} \ln Z_{N,p}(\beta; \xi)) \\ &= \ln 2 - (\alpha/2) \ln(1 - \beta) - (\alpha\beta^2)/(2(1 - \beta)) - (\alpha\beta/2), \end{aligned} \quad (25)$$

ξ -almost surely. The same limits hold for the quenched averages, so that in particular

$$\lim_{N \rightarrow \infty} N^{-1} \mathbb{E} \ln Z_{N,p}(\beta; \xi) = \ln 2 - \frac{\alpha}{2} \ln(1 - \beta) - \frac{\alpha\beta}{2},$$

where, in all these formulas, the last term, namely $-\alpha\beta/2$, arises due to the diagonal contribution of the complete partition function (4).

Theorem 3. *There exists $\beta_c(\alpha)$, defined by eq. (21), such that for $\beta < \beta_c(\alpha)$ we have the following convergence in distribution*

$$\ln \tilde{Z}_{N,p}(\beta; \xi) - \ln \mathbb{E} \tilde{Z}_{N,p}(\beta; \xi) \rightarrow C(\beta) + \chi S(\beta) \quad (26)$$

where χ is a unit Gaussian in $\mathcal{N}[0, 1]$ and

$$C(\beta) = -\frac{1}{2} \ln \sqrt{1/(1 - \sigma^2 \beta^2 \alpha)} \quad (27)$$

$$S(\beta) = \left(\ln \sqrt{1/(1 - \sigma^2 \beta^2 \alpha)} \right)^{\frac{1}{2}}, \quad (28)$$

with $\sigma = (1 - \beta)^{-1}$.

3 Extension to the replica symmetric solution

Once the correct interpolating structure is understood, and spurred by the observation that the replica symmetric expression for the quenched free energy of the three models, namely the analogical neural network, the SK spin glass and the Gaussian one, are well known and investigated (for instance in [30][7][26][9][11]) we want to push further the equivalence among neural network and spin glasses, giving a complete picture also of the replica symmetric approximation.

To this task, let us recall that the replica symmetric approximation of the quenched free energy of the analogical neural network $A_{NN}^{RS}(\alpha, \beta)$ is given by the following expression [9]

$$\begin{aligned} A_{NN}^{RS}(\alpha, \beta) &= \ln 2 + \int d\mu(z) \ln \cosh(z\sqrt{\alpha\beta\bar{p}}) + \frac{\alpha}{2} \ln\left(\frac{1}{1 - \beta(1 - \bar{q})}\right) + \\ &+ \frac{\alpha\beta}{2} \frac{\bar{q}}{1 - \beta(1 - \bar{q})} - \frac{\alpha\beta}{2} \bar{p}(1 - \bar{q}), \end{aligned} \quad (29)$$

where the order parameters denoted with a bar (to mean their RS approximation) are given by

$$\bar{q} = \int d\mu(z) \tanh^2(z\sqrt{\alpha\beta\bar{p}}), \quad (30)$$

$$\bar{p} = \beta\bar{q}/\left(1 - \beta(1 - \bar{q})\right)^2. \quad (31)$$

Let us introduce further β_1 and β_2 as

$$\beta_1 = \frac{\sqrt{\alpha}\beta}{1 - \beta(1 - \bar{q})}, \quad (32)$$

$$\beta_2 = 1 - \beta(1 - \bar{q}), \quad (33)$$

such that $\beta_1\beta_2 = \sqrt{\alpha}\beta$. We need also the RS approximation $A_{SK}^{RS}(\beta_1)$ of the quenched free energy of the SK model, at the noise level β_1 , namely

$$A_{SK}^{RS}(\beta_1) = \ln 2 + \int d\mu(z) \ln \cosh(\beta_1 \sqrt{\bar{q}_{SK}} z) + \frac{1}{4}\beta_1^2(1 - \bar{q}_{SK})^2, \quad (34)$$

where

$$\bar{q}_{SK} = \int d\mu(z) \tanh^2(\beta_1 z \sqrt{\bar{q}_{SK}}). \quad (35)$$

By a direct comparison among the overlap expressions (30, 35) we immediately conclude that we must have

$$\beta_1^2 \bar{q}_{SK} = \alpha \beta \bar{p},$$

which indeed holds as it can be verified easily, bearing in mind the expression (31) and (32) for \bar{p} and β_1 .

As a last ingredient we need to introduce also the replica symmetric expression $A_{Gauss}^{RS}(\beta_2, \beta)$ of the Gaussian spin glass at a noise level β_2 as [11]

$$A_{Gauss}^{RS}(\beta_2, \beta) = \frac{1}{2} \ln \sigma + \frac{1}{2}\beta_2^2 \bar{p}_G \sigma^2 + \frac{1}{4}\beta_2^2 \bar{p}_G^2, \quad (36)$$

where

$$\bar{p}_G = (\beta_2 - (1 - \beta))/\beta_2^2, \quad (37)$$

$$\sigma^2 = 1/(1 - \beta + \beta^2 \bar{p}_G). \quad (38)$$

Note that the definition of the overlap between continuous variables encoded by eq. (31) is in perfect agreement with the same overlap defined within the framework of eq.(37), because, being $\beta_2 = 1 - \beta(1 - \bar{q})$, we can write

$$\bar{p}_{Gauss} = \frac{\beta_2 - (1 - \beta)}{\beta_2^2} = \frac{1 - \beta(1 - \bar{q}) - (1 - \beta)}{(1 - \beta(1 - \bar{q}))^2} = \frac{\beta \bar{q}}{(1 - \beta(1 - \bar{q}))^2}. \quad (39)$$

As a consequence, through a direct verification by comparison (that we omit as it is long and straightforward), we can state the final theorem of the paper:

Theorem 4. *Fixed, at noise level β , β_1 and β_2 as in (32) and (33), the replica symmetric approximation of the quenched free energy of the analogical neural network can be linearly decomposed in terms of the replica symmetric approximation of the Sherrington-Kirkpatrick quenched free energy, at noise level β_1 , and the replica symmetric approximation of the quenched free energy of the Gaussian spin glass, at noise level β_2 , such that*

$$A_{NN}^{RS}(\beta) = A_{SK}^{RS}(\beta_1) - \frac{1}{4}\beta_1^2 + \alpha A_{Gauss}(\beta_2, \beta), \quad (40)$$

and the inequality (21) becomes an identity for the RS behavior.

Remark 2. *We stress that the above Theorem is in agreement with the sum rule (20) of Section 2 as, in the replica symmetric approximation, $q_{\sigma\sigma'} = \bar{q}$ and $p_{zz'} = \bar{p}$, hence*

$$\beta_1 \bar{q} - \sqrt{\alpha} \beta_2 \bar{p} = \frac{\sqrt{\alpha} \beta \bar{q}}{(1 - \beta(1 - \bar{q}))^2} - \sqrt{\alpha} (1 - \beta(1 - \bar{q})) \frac{\beta \bar{q}}{(1 - \beta(1 - \bar{q}))^2} = 0. \quad (41)$$

Remark 3. *Approaching the high-temperature region we have $\bar{q} \rightarrow 0$ and $\bar{p} \rightarrow 0$, and clearly $\beta \rightarrow 1/(1 + \sqrt{\alpha})$. As a consequence we have*

$$\beta_2 = 1 - \beta(1 - \bar{q}) \rightarrow 1 - 1/(1 + \sqrt{\alpha}), \quad (42)$$

$$\beta_1 = \frac{\sqrt{\alpha} \beta}{1 - \beta(1 - \bar{q})} \rightarrow 1, \quad (43)$$

then $\beta + \beta_2 \rightarrow 1$, such that also the single-party counterparts approaches their critical points.

Coherently, inside the annealed region we get $\bar{q} = 0$, then with the expressions for β_1, β_2 we can write $\beta_2 + \beta = 1$ that is the boundary of the annealed region for the Gaussian spin glass, while $\beta_1 = \sqrt{\alpha} \beta / (1 - \beta)$ because $\beta \leq 1/(1 + \sqrt{\alpha})$ we get $\beta_1 \leq 1$, namely the annealed region of the SK model.

4 Conclusions and Outlook

Neural networks are becoming the paradigm of a wide family of complex systems with *cognitive capabilities* such as memory and learning both in the living world and outside.

As a consequence, a solid control of these networks is fundamental: In this paper we provided a clear analysis of the analogical neural network thought

of as a bipartite spin glass, made of by two different type of spins: one ensemble of dichotomic variables, as in the celebrated Sherrington-Kirkpatrick model, and one ensemble made of by Gaussian distributed variables.

Exploiting this analogy, we developed a new interpolation scheme among the bipartite spin glass that mirrors the neural network and two independent glassy systems. Through this novel technique, we have then shown how to get a complete control of the annealed region of the neural network: The critical line has been obtained, together with an explicit behavior of all the main thermodynamical quantities: free energy, internal energy, entropy and overlaps (namely the order parameters of the theory).

One step forward we extended our interpolation scheme beyond the ergodic region, under the assumption of replica symmetry: We showed that the replica symmetric approximation of the quenched free energy of the analogical Hopfield model (at noise level β) can be expressed in terms of the replica symmetric expressions of the quenched free energies of the SK model (at noise level β_1) and of the Gaussian model (at noise level β_2), and we obtained the equations linking β, β_1, β_2 obtaining then a complete control also within this framework.

All that opens very interesting perspectives. The structure of the neural network as a linear combination of spin glasses is very rich: in fact we know that, as the SK model presents a very glassy full RSB structure [28], in the Gaussian one this is absent, since the true solution is in fact RS even with no external field [11]. Thus one could aspect in our analogical neural network a competition of these two effects: rather a new feature in the complex systems scenario, that has to be deeply investigated.

Clearly we would deepen this topic, for example within a fully broken replica symmetry scenario on which we plan to report soon.

Furthermore, the analogical model shares many features with the original Hopfield model (which is even harder from a mathematical point of view) for which one could study in what measure this structure is preserved.

Future outlooks should cover also the completely analogical model in order to develop mathematical techniques beyond the standard ones required in artificial intelligence and closer to system biology.

Acknowledgements

The strategy of this work is founded by the MIUR trough the FIRB grant RBFR08EKEV which is acknowledged, together with Sapienza Università di Roma for partial financial support. Partial support from INFN is also acknowledged.

Appendix.

Fluctuation Theory for the Order Parameters

We develop in this appendix a fluctuation theory of the order parameters to see that the ergodicity breaking is accomplished through a second order phase transition (i.e. the overlap fluctuations, properly rescaled over the volume, do diverge on the line $\beta_c(\alpha)$ hence defining a critical phenomenon). To satisfy this task we proceed as follows: at first we introduce a different interpolating structure with respect to the one discussed above (developed and discussed in [10]) to bridge the neural network with two single party one-body models where spins are subjected to random fields in a way close to stochastic stability [40] or cavity perspective [29]. Then we evaluate the flow with respect to the interpolating parameter so to be able to calculate variations of generic observable as overlap correlation functions.

Then we define the centered and rescaled overlaps and introduce their correlation matrix. Each element of this matrix then is evaluated at $t = 0$ and then propagated through $t = 1$ via its flow: This procedure encodes naturally for a system of coupled linear differential equations that, once solved, gives the expressions of the overlap fluctuations. The latter are found to diverge on the critical line $\beta_c(\alpha)$ already outlined and this will close our inspection of the annealed regime.

Let us start the plan by introducing the next interpolating structure:

In a pure stochastic stability fashion [10], we need to introduce also two classes of i.i.d. $\mathcal{N}[0, 1]$ variables, namely N variables η_i and p variables $\tilde{\eta}_\mu$, whose average is still encoded into the \mathbb{E} operator and by which we define the following interpolating quenched pressure $\tilde{\varphi}_{N,p}(\beta, t)$

$$\begin{aligned} \tilde{\varphi}_{N,p}(\beta, t) = & \frac{1}{N} \mathbb{E} \log \sum_{\sigma} \int \prod_{\mu}^p d\mu(z_{\mu}) \exp(\sqrt{t} \sqrt{\frac{\beta}{N}} \sum_{i,\mu} \xi_i^{\mu} \sigma_i z_{\mu}) \quad (44) \\ & \cdot \exp(a\sqrt{1-t} \sum_i \eta_i \sigma_i) \exp(b\sqrt{1-t} \sum_{\mu} \tilde{\eta}_{\mu} z_{\mu}) \exp(c \frac{(1-t)}{2} \sum_{\mu} z_{\mu}^2), \end{aligned}$$

where

$$a = \sqrt{\alpha\beta\bar{p}}, \quad b = \sqrt{\beta\bar{q}} \quad c = \beta(1 - \bar{q}).$$

We stress that $t \in [0, 1]$ interpolates between $t = 0$ where the interpolating quenched pressure becomes made of by non-interacting systems (a series of one-body problem) whose integration is straightforward (as well as the evaluation of the overlap correlation functions it produces) and the opposite limit, $t = 1$, that recovers the correct quenched free energy. Then we can evaluate

the flow with respect to the Boltzman factor encoded in the structure (44) as stated in the next

Proposition A. *Given O as a smooth function of s replica overlaps (q_1, \dots, q_s) and (p_1, \dots, p_s) , the following streaming equation holds:*

$$\begin{aligned} \frac{d}{dt}\langle O \rangle_t &= \beta\sqrt{\alpha}\left(\sum_{a,b}^s \langle O \cdot \xi_{a,b}\eta_{a,b} \rangle_t \right. \\ &\quad \left. - s \sum_{a=1}^s \langle O \cdot \xi_{a,s+1}\eta_{a,s+1} \rangle_t + \frac{s(s+1)}{2} \langle O \cdot \xi_{s+1,s+2}\eta_{s+1,s+2} \rangle_t \right). \end{aligned} \quad (45)$$

We skip the proof as it is long but simple and works by a direct evaluation which is pretty standard in the disordered system literature (see for example [30, 7]).

The rescaled overlap ξ_{12} and η_{12} are defined accordingly to

$$\xi_{12} = \sqrt{N}\left(q_{12} - \bar{q}\right), \quad (46)$$

$$\eta_{12} = \sqrt{K}\left(p_{12} - \bar{p}\right). \quad (47)$$

In order to control the overlap fluctuations, namely $\langle \xi_{12}^2 \rangle_{t=1}$, $\langle \xi_{12}\eta_{12} \rangle_{t=1}$, $\langle \eta_{12}^2 \rangle_{t=1}$, ..., noting that the streaming equation pastes two replicas to the ones already involved ($s = 2$ so far), we need to study nine correlation functions. It is then useful to introduce them and link them to capital letters so to simplify their visualization:

$$\langle \xi_{12}^2 \rangle_t = A(t), \quad \langle \xi_{12}\xi_{13} \rangle_t = B(t), \quad \langle \xi_{12}\xi_{34} \rangle_t = C(t), \quad (48)$$

$$\langle \xi_{12}\eta_{12} \rangle_t = D(t), \quad \langle \xi_{12}\eta_{13} \rangle_t = E(t), \quad \langle \xi_{12}\eta_{34} \rangle_t = F(t), \quad (49)$$

$$\langle \eta_{12}\eta_{12} \rangle_t = G(t), \quad \langle \eta_{12}\eta_{13} \rangle_t = H(t), \quad \langle \eta_{12}\eta_{34} \rangle_t = I(t). \quad (50)$$

If we introduce the operator *dot* as

$$\dot{O} = \frac{1}{\beta\sqrt{\alpha}} \frac{dO}{dt},$$

which simplifies calculations and shifts the propagation of the flow from $t = 1$ to $t = \beta\sqrt{\alpha}$. Assuming a Gaussian behavior, as in the strategy outlined in

[30], we can write the overall flow of the overlap correlation functions in the form of the following differential system

$$\begin{aligned}
\dot{A} &= 2AD - 8BE + 6CF, \\
\dot{B} &= 2AE + 2BD - 4BE - 6BF - 6EC + 12CF, \\
\dot{C} &= 2AF + 2CD + 8BE - 16BF - 16CE + 20CF, \\
\dot{D} &= AG - 4BH + 3CI + D^2 - 4E^2 + 3F^2, \\
\dot{E} &= AH + BG - 2BH - 3BI - 3CH + 6CI + 2ED - 2E^2 - 6EF + 6F^2, \\
\dot{F} &= AI + CG + 4BH - 8BI - 8CH + 10CI + 2DF + 4E^2 - 16EF + 10F^2, \\
\dot{G} &= 2GD - 8HE + 6IF, \\
\dot{H} &= 2GE + 2HD - 4HE - 6HF - 6IE + 12IF, \\
\dot{I} &= 2GF + 2DI + 8HE - 16HF - 16IE + 20IF.
\end{aligned}$$

Although it may appear complex, it is relatively easy to solve this system, once the initial conditions at $t = 0$ are known (information then can be obtained straightforwardly as at $t = 0$ everything factorizes the theory being one-body). Our general analysis covers also the case where external fields are involved. We do not report here the full analysis, for the sake of brevity. Here, as we are interested in finding where ergodicity becomes broken, we start propagating $t \in 0 \rightarrow 1$ from the annealed region, where $\bar{q} = 0$ and $\bar{p} = 0$, which simplifies further the problem:

In fact, it is immediate to check that, for the only terms that we need to consider, A, D, G (the other being strictly zero on the whole $t \in [0, 1]$), the starting points are $A(0) = 1, D(0) = 0, G(0) = (1 - \beta)^{-2}$ and their evolution is ruled by

$$\dot{A} = 2AD, \tag{51}$$

$$\dot{D} = AG + D^2, \tag{52}$$

$$\dot{G} = 2GD. \tag{53}$$

The solution of this differential system is long but straightforward then we skip the proof and directly state the next

Theorem A. *In the ergodic region the behavior of the overlap fluctuations*

is regular and described by the following equations

$$\langle \xi_{12}^2 \rangle = \frac{(1 - \beta)^2}{(1 - \beta)^2 - \beta^2 \alpha}, \quad (54)$$

$$\langle \xi_{12} \eta_{12} \rangle = \frac{\beta \sqrt{\alpha}}{(1 - \beta)^2 - \beta^2 \alpha}, \quad (55)$$

$$\langle \eta_{12}^2 \rangle = \frac{1}{(1 - \beta)^2 - \beta^2 \alpha}, \quad (56)$$

diverging on the critical line $\beta_c(\alpha)$, defined by eq. (21), hence defining a second order phase transition.

References

- [1] J.A. Acebron, L. Bonilla, C. Perez-Vicente, Prez; F. Ritort, R. Spigler, *The Kuramoto model: a simple paradigm for synchronization phenomena*, Rev. Mod. Phys. **77**, 137, (2005).
- [2] E. Agliari, A. Barra, F. Guerra, F. Moauro, *A thermodynamical perspective of immune capabilities*, J. Theor. Biol. **287**, 48, 2011.
- [3] M. Aizenman, J. Lebowitz, D. Ruelle, *Some Rigorous Results on the Sherrington-Kirkpatrick Model of Spin Glasses*, Commun. Math. Phys., **112** 3-20 (1987).
- [4] S. Albeverio, B. Tirozzi, B. Zegarlinski *Rigorous results for the free energy in the Hopfield model*, Comm. Math. Phys. **150**, 337 (1992).
- [5] D.J. Amit, *Modeling brain function: The world of attractor neural network*, Cambridge University Press, (1992).
- [6] D.J. Amit, H. Gutfreund, H. Sompolinsky *Storing infinite numbers of patterns in a spin glass model of neural networks*, Phys. Rev. Lett. **55**, 1530-1533, (1985).
- [7] A. Barra, *Irreducible free energy expansion and overlap locking in mean field spin glasses*, J. Stat. Phys. **123**, 601-614 (2006).
- [8] A. Barra, P. Contucci, *Toward a quantitative approach to migrants integration*, Europhys. Lett. **89**, 68001, (2010).
- [9] A. Barra, F. Guerra, *About the ergodic regime in the analogical Hopfield neural networks. Moments of the partition function*, J. Math. Phys. **49**, 125217 (2008).

- [10] A. Barra, G. Genovese, F. Guerra, *The replica symmetric behavior of the analogical neural network*, J. Stat. Phys. **140**, 784, (2010).
- [11] A. Barra, G. Genovese, F. Guerra, D. Tantari, *A solvable mean field model of a Gaussian spin glass*, submitted, available at arXiv:1109.4069.
- [12] A.Barra, G.Genovese, F.Guerra, *Equilibrium statistical mechanics of bipartite spin systems*, J. Phys. A: Math. and Theor. **44**, 245002, (2011).
- [13] J. P. Bigus, *Data Mining With Neural Networks: Solving Business Problems from Application Development to Decision Support*, McGraw-Hill (2006).
- [14] D. Bollé, T. M. Nieuwenhuizen, I. Perez-Castillo, T. Verbeiren, *A spherical Hopfield model*, J. Phys. A **36**, 10269, (2003).
- [15] A. Bovier, A.C.D. van Enter and B. Niederhauser, *Stochastic symmetry-breaking in a Gaussian Hopfield-model*, J. Stat. Phys. **95**, 181-213 (1999).
- [16] A. Bovier, V. Gayrard *An almost sure central limit theorem for the Hopfield model*, Markov Proc. Rel. Fields **3**, 151-173 (1997).
- [17] A. Bovier *Self-averaging in a class of generalized Hopfield models*, J. Phys. A **27**, 7069-7077 (1994).
- [18] A. N. Burkitt, *A review of the integrate-and-fire neuron model*, Biol Cybern **95**, 1, (2006).
- [19] A.C.C. Coolen, *The mathematical theory of minority games: Statistical mechanics of interacting agents*, Oxford University Press, (2005).
- [20] A.C.C. Coolen, A.J. Noest, G.B. de Vries, *Modelling Chemical Modulation of Neural Processes*, Network **4**, 101, (1993).
- [21] A.C.C. Coolen, R. Kuehn, P. Sollich, *Theory of Neural Information Processing Systems*, Oxford University Press, 2005.
- [22] D. De Martino, M. Figliuzzi, A. De Martino, E. Marinari, *Computing fluxes and chemical potential distributions in biochemical networks: energy balance analysis of the human red blood cell*, submitted. Available at arXiv:1107.2330 (2012).
- [23] A. Di Biasio, E. Agliari, A. Barra, R. Burioni, *Cooperativity in chemical kinetics*, Theor. Chem. Acc. **131**, 1104, (2012).

- [24] R. Dietrich, M. Oppen, and H. Sompolinsky. *Statistical mechanics of support vector networks*, Phys. rev. lett. **82**, 2975, (1999).
- [25] R.S. Ellis, *Large deviations and statistical mechanics*, Springer, New York, 1985.
- [26] G. Genovese, A. Barra, *A mechanical approach to mean field spin models*, J. Math. Phys. **50**, 365234 (2009).
- [27] F. Guerra, *An introduction to mean field spin glass theory: methods and results*, In: *Mathematical Statistical Physics*, A. Bovier et al. eds, 243 – 271, Elsevier, Oxford, Amsterdam, 2006.
- [28] F. Guerra, *Broken Replica Symmetry Bounds in the Mean Field Spin Glass Model*, Commun. Math. Phys. **233:1**, 1-12 (2003).
- [29] F. Guerra, *About the overlap distribution in mean field spin glass models*, Int. Jou. Mod. Phys. B **10**, 1675-1684 (1996).
- [30] F. Guerra, *Sum rules for the free energy in the mean field spin glass model*, in *Mathematical Physics in Mathematics and Physics: Quantum and Operator Algebraic Aspects*, Fields Institute Communications **30**, American Mathematical Society (2001).
- [31] F. Guerra, F. L. Toninelli, *The Thermodynamic Limit in Mean Field Spin Glass Models*, Commun. Math. Phys. **230:1**, 71-79 (2002).
- [32] D.O. Hebb, *Organization of Behaviour*, Wiley, New York, 1949.
- [33] J.J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Ntl. Acad. Sci. USA **79**, 2554-2558 (1982).
- [34] M. Mézard, A. Montanari, *Information, Physics and Computation*, Oxford University press, 2009.
- [35] M. Mézard, G. Parisi and M. A. Virasoro, *Spin glass theory and beyond*, World Scientific, Singapore, 1987.
- [36] L. Pastur, M. Scherbina, B. Tirozzi, *The replica symmetric solution of the Hopfield model without replica trick* J. Stat. Phys. **74**, 1161-1183 (1994).
- [37] L. Pastur, M. Scherbina, B. Tirozzi, *On the replica symmetric equations for the Hopfield model* J. Math. Phys. **40**, 3930-3947 (1999).

- [38] I. Perez-Castillo, B. Wemmenhove, J.P.L. Hatchett, A.C.C. Coolen, N.S. Skantzos and T. Nikolettopoulos, *Analytic solution of attractor neural networks on scale-free graphs*, J. Phys. A **37**, 8789, (2004).
- [39] Y. Singh, A.S. Chauhan, *Neural networks in data mining*, J. Theor. and Appl. Inform. Techn. **5**, 1, (2009).
- [40] P. Sollich, A. Barra, *Notes on the polynomial identities in random overlap structures*, J. Stat. Phys. **147**, 351, (2012).
- [41] M. Talagrand, *Spin glasses: a challenge for mathematicians. Cavity and mean field models*, Springer-Verlag, (2003).
- [42] M. Talagrand, *Rigorous results for the Hopfield model with many patterns*, Probab. Th. Relat. Fields **110**, 177-276 (1998).
- [43] M. Talagrand, *Exponential inequalities and convergence of moments in the replica-symmetric regime of the Hopfield model*, Ann. Probab. **38**, 1393-1469 (2000).
- [44] B. Wemmenhove, A.C.C. Coolen, *Finite connectivity attractor neural networks*, J. Phys. A: Math. and Gen. **36**, 9617, (2003).